4F7 Adaptive Filters (and Spectrum Estimation)

Estimation for Hidden Markov Models

Sumeetpal Singh

Engineering Department

Email : sss40@eng.cam.ac.uk

Course website:

www-sigproc.eng.cam.ac.uk/~sss40/teaching.html

1 Motivating Example

At a casino a fair die is used but occasionally switch to a biased die

The fair die has prob. 1/6 for each number turning up but the biased die has

(outcome,prob)={(1,0.1),(2,0.1),(3,0.1),(4,0.1),(5,0.1),(6,0.5)}

After each roll the next die to be used is selected with probabilities:
Prob(next=fair|current=fair)=0.95,
Prob(next=biased|current=fair)=0.05,
Prob(next=fair|current=biased)=0.1,
Prob(next=biased|current=biased)=0.9,

Here are example outcomes of 20 throws of the fair die
1 5 5 5 5 3 2 1 3 4 2 4 2 1 3 5 5 1 1 5
and the unfair die
6 6 4 4 6 2 6 6 6 6 5 6 6 6 6 3 2 6 6 6

Problem: given the outcomes from throws 1 to T, how do we evaluate the probability of cheating?

2 Definition of a Hidden Markov Model

1. Set of states: $S = \{1, 2, \ldots, n\}$

2. Set of observations: $O = \{1, 2, \ldots, m\}$

3. State transition probability matrix P with $[P]_{i,j} = p_{i,j} = \text{Pr}(\text{next state } j | \text{ current state } i)$

4. Observation probability matrix Q with $[Q]_{i,j} = q_{i,j} = \text{Pr}(\text{of getting obs. } j \text{ in state } i)$

5. Initial state distribution at time 0: $\pi_0 = (\pi_0(1), \pi_0(2), \ldots, \pi_0(n))$

The HMM is now completely specified given ingredients 1 to 5

Main points: The hidden state process $\{x_t\}_{t=0}^{t=T}$ is a Markov chain. We don't observe the realization of the hidden state process directly but do so via an observation process $\{y_t\}_{t=1}^{t=T}$

We would like to perform the following tasks ...

Filtering: compute $\pi_t(x_t) = \Pr(x_t|y_{1:t})$ at time t recursively where $y_{1:t}$ denotes the set of observations $\{y_1, y_2, \dots, y_t\}$

Smoothing: given $\{y_1, y_2, \dots, y_T\}$ compute $\Pr(x_t|y_{1:T})$ for all $t = 0, 1, \dots, T$. This is solved by the *forward-backward* algorithm

Maximum a posteriori (MAP) estimate

$$x^*_{0:T} = \arg\max_{x_{0:T}} \Pr(x_{0:T}|y_{1:T})$$

This is solved by the Viterbi algorithm

3 The Law of the HMM

The probability of getting hidden states $x_{0:T}$ and observing $y_{1:T}$ is

$$\Pr(x_{0:T}, y_{1:T})$$
$$= \pi_0(x_0) p_{x_0,x_1} q_{x_1,y_1} p_{x_1,x_2} q_{x_2,y_2} \cdots p_{x_{T-1},x_T} q_{x_T,y_T}$$

We write this expression using the fact that the hidden state process is Markov chain

$$\Pr(x_{0:T}) = \pi_0(x_0) p_{x_0,x_1} p_{x_1,x_2} \cdots p_{x_{T-1},x_T}$$

and the observation probability $\Pr(y_{1:T}|x_{0:T})$ factors as

$$\prod_{i=1}^{T} \Pr(y_i|x_i) = \prod_{i=1}^{T} q_{x_i,y_i}$$

4 Filtering

To solve filtering problem, let $\pi_t(i) = \Pr(x_t = i | y_{1:t})$. There are two main steps. The first is the *prediction* step

$$\text{Prediction:} \;\; \Pr(x_{t+1} | y_{1:t}) = \sum_{x_t} p_{x_t, x_{t+1}} \pi_t(x_t)$$

The second step is the update step

$$\text{Update:} \;\; \pi_{t+1}(x_{t+1}) = \frac{q_{x_{t+1}, y_{t+1}} \Pr(x_{t+1} | y_{1:t})}{\sum_{x_{t+1}} q_{x_{t+1}, y_{t+1}} \Pr(x_{t+1} | y_{1:t})}$$

We can combine both steps and write it in matrix form. Regard π_t as the vector $[\pi_t(1), \pi_t(2), \ldots, \pi_t(n)]^\top$ and let $B(y_{t+1})$ be the diagonal matrix

$$B(y_{t+1}) = \begin{bmatrix} q_{1,y_{t+1}} & 0 & \cdots & 0 \\ 0 & q_{2,y_{t+1}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & q_{n,y_{t+1}} \end{bmatrix}$$

then

$$\pi_{t+1}^\top = \frac{\pi_t^\top P B(y_{t+1})}{\pi_t^\top P B(y_{t+1}) \mathbf{1}}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^\top$

5 Smoothing

To solve the smoothing problem, we need the following result

$$\begin{aligned}
&\Pr(y_{t+1}, y_{t+2}, \ldots, y_T | x_t) \\
&= \sum_{x_{t+1}=1}^{n} \Pr(y_{t+2}, y_{t+3}, \ldots, y_T | x_{t+1}) q_{x_{t+1}, y_{t+1}} p_{x_t, x_{t+1}}
\end{aligned}$$

We derive this result as follows:

$$\begin{aligned}
\Pr(y_{t+1:T} | x_t) &= \sum_{x_{t+1}} \Pr(y_{t+1:T}, x_{t+1} | x_t) \\
&= \sum_{x_{t+1}} \underbrace{\Pr(y_{t+2:T}, | y_{t+1}, x_{t+1}, x_t)}_{\Pr(y_{t+2:T} | x_{t+1})} \\
&\times \underbrace{\Pr(y_{t+1} | x_{t+1}, x_t)}_{q_{x_{t+1}, y_{t+1}}} \underbrace{\Pr(x_{t+1} | x_t)}_{p_{x_t, x_{t+1}}}
\end{aligned}$$

We call $\beta_t(x_t) = \text{Pr}(y_{t+1}, y_{t+2}, \ldots, y_T | x_t)$ the *backward* recursion

It is computed starting at $T - 1$ in the following order $\beta_{T-1}, \beta_{T-2}, \ldots, \beta_0$

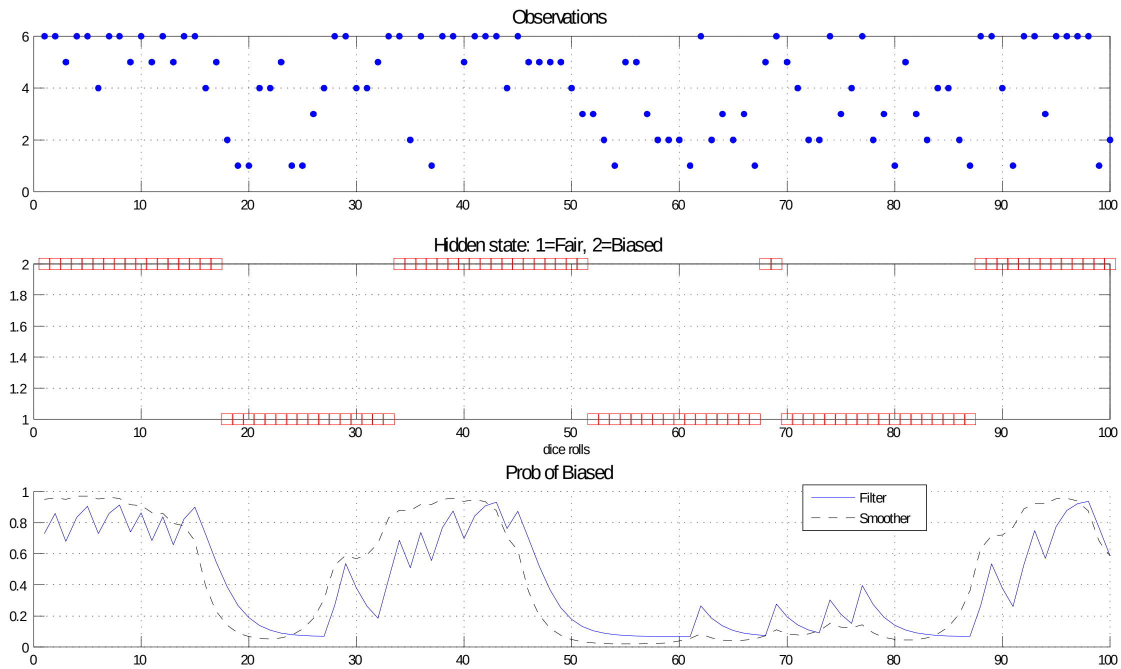
It admits a recursion similar to the filter π_t and can be expressed as

$$\beta_t = PB(y_{t+1})\beta_{t+1}$$

with $\beta_T = [1, \ldots, 1]^\mathsf{T}$ (initialized to the vector of ones)

Once we have computed β_t,

$$\text{Pr}(x_t | y_{1:T}) = \frac{\pi_t(x_t)\beta_t(x_t)}{\pi_t^\mathsf{T}\beta_t}$$



Additional Reading:

Rabiner, L.W., "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, 1989. (availabe on course website)